

RESEARCH ARTICLE

ThermoBase: A database of the phylogeny and physiology of thermophilic and hyperthermophilic organisms

Juliana DiGiacomo¹*, Christopher McKay¹, Alfonso Davila¹

NASA Ames Research Center, Moffett Field, California, United States of America

* These authors contributed equally to this work.

* juldigiaco@gmail.com

Abstract

Thermophiles and hyperthermophiles are those organisms which grow at high temperature (> 40°C). The unusual properties of these organisms have received interest in multiple fields of biological research, and have found applications in biotechnology, especially in industrial processes. However, there are few listings of thermophilic and hyperthermophilic organisms and their relevant environmental and physiological data. Such repositories can be used to standardize definitions of thermophile and hyperthermophile limits and tolerances and would mitigate the need for extracting organism data from diverse literature sources across multiple, sometimes loosely related, research fields. Therefore, we have developed ThermoBase, a web-based and freely available database which currently houses comprehensive descriptions for 1238 thermophilic or hyperthermophilic organisms. ThermoBase reports taxonomic, metabolic, environmental, experimental, and physiological information in addition to literature resources. This includes parameters such as coupling ions for chemiosmosis, optimal pH and range, optimal temperature and range, optimal pressure, and optimal salinity. The database interface allows for search features and sorting of parameters. As such, it is the goal of ThermoBase to facilitate and expedite hypothesis generation, literature research, and understanding relating to thermophiles and hyperthermophiles within the scientific community in an accessible and centralized repository. ThermoBase is freely available online at the Astrobiology Habitable Environments Database (AHED; <https://ahed.nasa.gov>), at the Database Center for Life Science (TogoDB; <http://togodb.org/db/thermobase>), and in the [S1 File](#).

OPEN ACCESS

Citation: DiGiacomo J, McKay C, Davila A (2022) ThermoBase: A database of the phylogeny and physiology of thermophilic and hyperthermophilic organisms. PLoS ONE 17(5): e0268253. <https://doi.org/10.1371/journal.pone.0268253>

Editor: William J. Brazelton, University of Utah, UNITED STATES

Received: October 5, 2021

Accepted: April 25, 2022

Published: May 10, 2022

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](#) public domain dedication.

Data Availability Statement: Thermobase is freely available at <http://togodb.org/db/thermobase>. The generated spreadsheet database in this study is included in the article and [Supplementary Information](#).

Funding: This work was supported by the NASA Internal Scientist Funding Model (Habitable Environments and Biosignatures Project). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The funder provided support in the form of salaries for authors A.D. and C.M. but did

Introduction

Thermophiles and hyperthermophiles constitute two sub-classes of extremophile microorganisms which grow optimally above high (40–50°C) and extremely high (70–80°C) temperatures, respectively [1]. Phylogenetic and physiological evidence suggests an extremely ancient origin for these organisms which parallels the origins of life and the last universal common ancestor (LUCA) [2, 3]. In both the Archaeal and Bacterial domains, phylogenetic analysis places most hyperthermophiles and a disproportionate number of thermophiles near their evolutionary

not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

Competing interests: The authors have declared that no competing interests exist.

nodes. Since the development of the earliest phylogenetic trees, which genetically discriminated between the three domains of life based upon the small subunit ribosomal RNA gene, it has been acknowledged that the earliest branches comprising the Bacteria and Archaea are populated by thermophiles [4]. This observation has been confirmed with further reconstructions of the tree of life [2, 5, 6].

The mechanisms of physiological and genetic adaptation to high temperature continues to be a subject of research and multiple factors have been proposed to contribute to high temperature tolerance, including membrane composition, membrane bioenergetics, or the thermal stability of enzymes, DNA and RNA [7]. Research into the origin and nature of thermophiles and hyperthermophiles also has applications for understanding the origin and evolution of life on Earth and the possibility of life on other planetary bodies- the study of Astrobiology. The ancient nature of thermophiles and hyperthermophiles aligns with theories that chemical disequilibria in hydrothermal vents could have provided a geochemical template for the evolution of the first metabolic networks [3]. Modern hydrothermal vents host rich ecosystems in which thermophiles and hyperthermophiles maintain the ability to generate thermodynamic disequilibrium through the reduction of locally available sulfur, iron, and carbon dioxide [3]. Sulfur and iron are of particular interest considering these species would have been water soluble when Earth's atmosphere was anoxic at the time of the hypothesized origin of life [3]. More recently, thermophiles and hyperthermophiles have become a biological analog for possible forms of life in extraterrestrial environments, especially the subsurface oceans of the icy moons of the outer solar system such as Europa or Enceladus. Hydrothermal activity similar to that found in Earth's oceans appears to also be prevalent in these Ocean Worlds, representing an engine of chemical disequilibrium and a source of redox energy and nutrients that life could exploit, akin to terrestrial thermophilic methanogenic archaea, which currently hold the record for growth at the highest temperature [8].

Thermophiles possess biological adaptations which allow the organism and its component parts to function and remain stable in otherwise inhospitable environments. In biotechnology, crucial methods such as Polymerase Chain Reaction (PCR) require the functionalities of polymerase from *Pyrococcus furiosus* (Pfu-Pol) or *Thermophilus aquaticus* (Taq-Pol) [9]. Such thermotolerant polymerases significantly increased the efficiency of this DNA replication method as it allowed for multiple thermal cycles without the need for constant replenishing of the denatured enzyme. RNA and DNA ligases from such heat tolerant organisms also have applications in such integral methods as Gibson assembly (Tag-ligase), Ligase Chain Reaction, and construction of mRNA sequencing libraries where enzymes not only function at high temperature but are also able to ligate combinations of blunt and cohesive ends [9].

Industrial processes have also capitalized on the ability for enzymes from thermophiles and hyperthermophiles to function at high temperatures. The sugar industry, paper industry, fruit industry, starch processing, food processing, alcohol production, lactose-free milk production, animal feed production, and laundry detergents are just a fraction of the industrial applications for thermostable enzymes [10]. Such enzymes pose several advantages over their mesophilic counterparts where efficacy at high temperatures reduces the risk of contamination and viscosity of the reaction medium, while increasing bioavailability and solubility of organic compounds, and the diffusion coefficient and concentrations of substrates and products allowing for high reaction rates [11]. The production and diversity of these thermostable enzymes has continued to grow in recent years with advances in isolation methods of thermophilic Archaea, Bacteria, and Eukarya from different ecological niches [11].

Outside of isolated thermostable enzymes, the metabolic activities of thermophiles and hyperthermophiles also have applications in biotechnology and industrial processes. Directly, the ancient metabolisms which are utilized by most thermophilic archaea have been re-

directed to produce simple C1 compounds such as carbon monoxide, formate, carbon dioxide, and methane, molecular hydrogen, and reduced/oxidized sulfur or transition metals [9]. However, these organisms also have metabolic applications for the bioconversion of xylose to ethanol (*Thermoanaerobacter ethanolicus*), crude oil degradation (*Bacillus sp.*), heavy metal recovery (*Bacillus sp.*), keratin degradation (*Fervidobacterium pennavorans*), saccharification of agricultural residues (*Sporotrichum thermophile*), hygiene indication in dairy products (*Anoxybacillus flavithermus*), remediation of textile dyes (*Geobacillus thermocatenulatus*), and even breast cancer treatments (*Aspergillus terreus*) [10].

The efficiency of scientific and technological investigation benefits from access to consolidated databases of relevant organism information. However, the generation of a database of thermophiles is especially difficult, rooted in the fact that thermophiles and hyperthermophiles are difficult to identify in a broad search. Primarily, this is because there is no universal definition for thermophiles or hyperthermophiles. Various sources posit differing minimum optimal temperatures for thermophilic classification. This value can range from 40°C, 45°C, or 50°C [12–14]. The cut off for hyperthermophiles is similarly unclear, described as 70°C, 75°C, or 80°C [12–14]. Due to the ambiguous nature of these definitions, classification of an organism's identity as a mesophile, thermophile or hyperthermophile can oftentimes conflict in published investigations. This makes it difficult to even identify thermophiles or hyperthermophiles accurately without referencing to the original literature resources which characterize optimal growth temperature. Besides this linguistic hurdle, the task of collecting additional environmental and physiological data poses an even larger expenditure of labor and time.

Over the past 10 years, several repositories have been developed in order to centralize and enhance understandings of thermophiles in the context of other extremophiles [15–17]. TEM-PURA is an online database which reports extreme growth temperatures in Prokaryotes [15]. ExProtDB is a database of extremophilic proteins and their host organisms with 259 relevant thermophiles identified [16]. ExtremeDB is a now inaccessible repository which once housed the general characteristics and genomic information for 865 extremophiles including 310 thermophiles [17]. Our database builds on these efforts by including extensive physiological and environmental data for all currently identified thermophiles.

ThermoBase was developed as a centralized database to facilitate the retrieval of primary environmental and physiological data pertaining to thermophile and hyperthermophiles. This freely available repository currently hosts 1238 thermophiles and hyperthermophiles species from the three domains of life. ThermoBase reports taxonomic, metabolic, environmental, and physiological information in addition to literature resources. This includes parameters such as ions for chemiosmosis, optimal pH and range, optimal temperature and range, optimal pressure, and optimal salinity. We utilize a lower-limit definition for thermophilicity (average optimal growth temperature $\geq 40^\circ\text{C}$) to allow for custom nomenclature definition preferences to be applied by the user. Therefore, with this 40°C limit, we include all known organisms which could possibly be considered thermophilic. The database format also allows for search features and sorting of other qualitative parameters. It is the goal of ThermoBase to facilitate and expedite hypothesis generation, literature research, and understanding relating to thermophiles and hyperthermophiles within the scientific community in an accessible and centralized repository.

Methods

ThermoBase was constructed from an initial database sourced from the work of Campbell et al. [19]. This work distinguished between psychrophilic, mesophilic, thermophilic, and hyperthermophilic species from the 7079 finished genomes from the Joint Genome Institute for the purpose of classifying ferredoxin and flavodoxin proteins [18]. Some accessible

metadata, including average pH, average temperature, ecosystem, energy source, oxygen requirement, and taxonomic data, was available in certain cases using Joint Genome Institute metadata labels.

The original dataset was then filtered to only include species specified as thermophiles or hyperthermophiles. However, the exact quantitative limits for these labels were not specified and had to be independently confirmed. The original metadata was also largely incomplete (<150 relevant species) and did not include all species nor desired parameters. ThermoBase was then expanded through a deep literature search of thermophilic and hyperthermophilic archaea, bacteria, and Eukarya in order to incorporate all appropriate species and available metadata. This required reference to multiple sources per species in order to collect data for taxonomic information, genome ID, general ecosystem, environment, energy source, metabolism, ions for chemiosmosis, oxygen requirement, optimal pH and range, optimal temperature and range, pressure at optimal temperature, optimal pressure, and optimal salinity.

ThermoBase was made available in various formats. The original file is downloadable as an Excel Spreadsheet from the NASA Astrobiology Habitable Environments Database (AHED; <https://ahed.nasa.gov>) and in the [S1 File](#). ThermoBase is also available as an online database developed using TogoDB, a database hosting service by the Database Center for Life Science. The original database file (comma-separated values file) was uploaded to the hosting service and formatted for proper integration. The back-end data is stored at a SPARQL endpoint (<http://togodb.org/sparql/thermobase>). The user interface was then customized and reformatted using HTML (HyperText Mark-up Language).

Results and discussion

ThermoBase Version 1.0 hosts detailed physiological and environmental descriptions of 1238 thermophilic and hyperthermophilic species with optimal growth temperatures equal to or above 40°C. The database encompasses all three domains of life with 373 archaeal, 836 bacterial, and 28 eukaryal species from over 226 distinct taxonomic families. Parameters for species metadata include generic information such as taxonomic information, genome ID, general ecosystem, environment, energy source, metabolism, and oxygen requirement as well as more specific categorizations such as ions for chemiosmosis, optimal pH and range, optimal temperature and range, pressure at optimal temperature, optimal pressure, and optimal salinity. Each entry also includes a key source for further investigation of a particular species. The total compiling of the database referenced over 590 unique scientific studies from 1967 to 2020 in an extensive literature search over the course of 11 months. Although all parameters could not be populated for all species, ThermoBase version 1.0 represents the extent of peer reviewed data which was available as of March 2022.

ThermoBase as a web-based service also allows for the searching of key words or values and sorting of its extensive qualitative and quantitative species data. In its Microsoft Excel Worksheet format, standard spreadsheet processing tools allow for users to reorganize and group single or multiple parameters at a time using sorting, custom sorting, and sorting level functionalities. Basic or advanced data filters can also be applied singly or cumulatively to isolate and stratify certain parameters or search terms of interest. These tools allow the user to focus their navigation of this extensive database for more efficient hypothesis generation, literature research, and understanding relating to thermophiles and hyperthermophiles within the scientific community and beyond.

The data parameters described by ThermoBase have immediate applications toward several major inquiries within studies of the limits of life, particularly the investigation of the relationship between heat tolerance and other taxonomic, physiological, or environmental properties.

Initial analyses of the database confirm that heat tolerance is phylogenetically widespread and involves diverse biochemical and physiological adaptations, with H⁺/Na⁺ bioenergetics and metabolism playing a significant role [19]. In addition, when organisms sorted by their maximum temperature, it is interesting that the top 50 species are exclusively within the domain of Archaea—those that have a maximum growth temperature >95°C. Within this group strict anaerobes dominate but a few facultative anaerobes are present. While a methanogen is found to function at the highest temperature, within the top 50 hyperthermophiles, sulfur reduction is the most common metabolic strategy. This is interesting to note as by some metrics there is about the same amount of research on methanogen hyperthermophiles as on sulfur reducing hyperthermophiles—based on a simple web search on the terms “hyperthermophiles methanogens” and “hyperthermophiles sulfur reducers”. The structure of ThermoBase allows for such observations to be ascertained with the potential to spark new academic research.

ThermoBase is freely available online at the NASA Astrobiology Habitable Environments Database (AHED; <https://ahed.nasa.gov>) and at the Database Center for Life Science (TogoDB; <http://togodb.org/db/thermobase>), and in the [S1 File](#). The ThermoBase online database continues to be manually updated with new scientific publications and user contributions.

Supporting information

S1 File. ThermoBase_ver_1.0_2022. A local spreadsheet of the ThermoBase database which was updated as of March 2022.
(XLSX)

Acknowledgments

We would like to thank Dr. Joff Silbery and Dr. Ian Campbell for their contributions of the original data set in Campbell et al. (20), and two anonymous reviewers who helped improve the original manuscript.

Author Contributions

Conceptualization: Juliana DiGiacomo, Christopher McKay.

Data curation: Juliana DiGiacomo.

Formal analysis: Juliana DiGiacomo.

Funding acquisition: Christopher McKay, Alfonso Davila.

Investigation: Juliana DiGiacomo.

Methodology: Juliana DiGiacomo.

Project administration: Christopher McKay, Alfonso Davila.

Resources: Juliana DiGiacomo, Christopher McKay, Alfonso Davila.

Supervision: Christopher McKay, Alfonso Davila.

Validation: Christopher McKay.

Visualization: Juliana DiGiacomo.

Writing – original draft: Juliana DiGiacomo.

Writing – review & editing: Christopher McKay, Alfonso Davila.

References

1. Lusk BG. Thermophiles; or, the modern prometheus: The importance of extreme microorganisms for understanding and applying extracellular electron transfer. *Frontiers in Microbiology*. 2019;10. <https://doi.org/10.3389/fmicb.2019.00010> PMID: 30728810
2. Ciccarelli FD. Toward automatic reconstruction of a highly resolved tree of life. *Science*. 2006; 311(5765):1283–7. <https://doi.org/10.1126/science.1123061> PMID: 16513982
3. Bhattacharya D, Friedl T, Schmidt H. The phylogeny of thermophiles and hyperthermophiles and the three domains of life. *Enigmatic Microorganisms and Life in Extreme Environments*. 1999;291–304.
4. Giulio M. The universal ancestor was a thermophile or a hyperthermophile: Tests and further evidence. *Journal of Theoretical Biology*. 2003; 221(3):425–36. <https://doi.org/10.1006/jtbi.2003.3197> PMID: 12642117
5. Gaucher EA, Kratzer JT, Randall RN. Deep phylogeny—how a tree can help characterize early life on Earth. *Cold Spring Harbor Perspectives in Biology*. 2009; 2(1).
6. Wang Q, Cen Z, Zhao J. The survival mechanisms of thermophiles at high temperatures: An angle of OMICS. *Physiology*. 2015; 30(2):97–106. <https://doi.org/10.1152/physiol.00066.2013> PMID: 25729055
7. Martin W, Baross J, Kelley D, Russell MJ. Hydrothermal vents and the origin of life. *Nature Reviews Microbiology*. 2008; 6(11):805–14. <https://doi.org/10.1038/nrmicro1991> PMID: 18820700
8. Takai K, Nakamura K, Toki T, Tsunogai U, Miyazaki M, Miyazaki J, et al. Cell proliferation at 122 C and isotopically heavy CH₄ production by a hyperthermophilic methanogen under high-pressure cultivation. *Proceedings of the National Academy of Sciences*. 2008; 105(31):10949–54. <https://doi.org/10.1073/pnas.0712334105> PMID: 18664583
9. Straub CT, Counts JA, Nguyen DM, Wu C-H, Zeldes BM, Crosby JR, et al. Biotechnology of extremely thermophilic archaea. *FEMS Microbiology Reviews*. 2018; 42(5):543–78. <https://doi.org/10.1093/femsre/fuy012> PMID: 29945179
10. Mehta R, Singhal P, Singh H, Damle D, Sharma AK. Insight into thermophiles and their wide-spectrum applications. *3 Biotech*. 2016; 6(1). <https://doi.org/10.1007/s13205-016-0368-z> PMID: 28330151
11. Gomes E, de Souza AR, Orjuela GL, Da Silva R, de Oliveira TB, Rodrigues A. Applications and benefits of thermophilic microorganisms and their enzymes for industrial biotechnology. *Fungal Biology*. 2016;459–92.
12. Thermophile Berenguer J. Gargaud M, editors. *Encyclopedia of Astrobiology*. Springer, Berlin, Heidelberg; 2011 p. 131.
13. Deacon J. Thermophilic microorganisms. 1997. Available from: <http://archive.bio.ed.ac.uk/jdeacon/microbes/thermo.htm>
14. Thermophile Berenguer J. *Encyclopedia of Astrobiology*. 2011:1666–7.
15. Sato Y, Okano K, Kimura H, Honda K. Tempura: Database of growth temperatures of usual and rare prokaryotes. *Microbes and Environments*. 2020; 35(3). <https://doi.org/10.1264/jsme2.ME20074> PMID: 32727974
16. Patra S. Extremophile protein database. 2018. Available from: <http://www.exprotdb.com/>
17. Majhi MC, Behera AK, Kulshreshtha NM, Mahmooduzafar D, Kumar R, Kumar A. ExtremeDB: A unified web repository of extremophilic Archaea and bacteria. *PLoS ONE*. 2013; 8(5). <https://doi.org/10.1371/journal.pone.0063083> PMID: 23696792
18. Campbell IJ, Bennett GN, Silberg JJ. Evolutionary relationships between low potential ferredoxin and flavodoxin electron carriers. *Frontiers in Energy Research*. 2019;7. <https://doi.org/10.3389/fenrg.2019.00079> PMID: 32095484
19. Mulkidjanian AY, Dibrov P, Galperin MY. The past and present of sodium energetics: May the Sodium-Motive Force be with you. *Biochimica et Biophysica Acta*. 2008; 1777(7–8):985–92. <https://doi.org/10.1016/j.bbabi.2008.04.028> PMID: 18485887